

Surface Sulfur Detection via Remote Sensing and Onboard Classification

LUKAS MANDRAKE, UMAA REBBAPRAGADA, KIRI L. WAGSTAFF, DAVID THOMPSON, STEVE CHIEN, DANIEL TRAN, ROBERT T. PAPPALARDO, DAMHNAIT GLEESON, REBECCA CASTAÑO

Jet Propulsion Laboratory, California Institute of Technology

Orbital remote sensing provides a powerful way to efficiently survey targets such as the Earth and other planets and moons for features of interest. One such feature of astrobiological relevance is the presence of surface sulfur deposits. These deposits have been observed to be associated with microbial activity at the Borup Fiord glacial springs in Canada, a location that may provide an analogue to other icy environments such as Europa.

This paper evaluates automated classifiers for detecting sulfur in remote sensing observations by the hyperion spectrometer on the EO-1 spacecraft. We determined that a data-driven machine learning solution was needed because the sulfur could not be detected by simply matching observations to sulfur lab spectra. We also evaluated several methods (manual and automated) for identifying the most relevant attributes (spectral wavelengths) needed for successful sulfur detection. Our findings include 1) the Borup Fiord sulfur deposits were best modeled as containing two sub-populations: sulfur on ice and sulfur on rock; 2) as expected, classifiers using Gaussian kernels outperformed those based on linear kernels, and should be adopted when onboard computational constraints permit; and 3) Recursive Feature Elimination selected sensible and effective features for use in the computationally constrained environment onboard EO-1. This study helped guide the selection of algorithm parameters and configuration for the classification system currently operational on EO-1. Finally, we discuss implications for a similar onboard classification system for a future Europa orbiter.

Categories and Subject Descriptors: I.4.6 [Image Processing and Computer Vision]: Segmentation—*pixel classification*; I.4.9 [Image Processing and Computer Vision]: Applications; I.5.2 [Pattern Recognition]: Design Methodology—*pattern analysis; feature evaluation and selection*

General Terms: Algorithms, Performance

Additional Key Words and Phrases: remote sensing, support vector machines, feature selection

1. INTRODUCTION

Satellite-based remote sensing has opened vast swaths of remote terrain to regular scientific scrutiny. Of particular interest are deserts, ice sheets, mountain ranges, and other nearly or totally inaccessible locales [Rango et al. 1989; Tucker et al. 1991; Goldstein et al. 1993]. These studies and others yield invaluable data on natural and human-induced climate change, seasonal variation, global-scale dynamics, and

Author's address: Lukas Mandrake, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Mail Stop 306-463, Pasadena, CA 91109, lukas.mandrake@jpl.nasa.gov.

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 0004-5411/20YY/0100-0001 \$5.00

a host of other relevant inquiries. Missions to other planets employ remote sensing to study targets too distant for direct human observation and analysis. However, the volume of data generated by Earth and remote spacecraft continues to exceed the available downlink bandwidth.

Several approaches have been employed to address this limitation. Data compression enabled the rescue of the science return of the Galileo spacecraft (which experienced a high gain antenna deployment failure) [Cheung et al. 1993] and continues to be a mainstay of satellite communication. Cooperation between in-situ spacecraft to relay data over the horizon or to boost transmission bandwidth has now also become possible [Hastrup et al. 2003] and is regularly employed for surface operations such as those of the Mars Exploration Rovers. However, the development of new sensors continues to outpace increases in data throughput capability. Spacecraft now carry hyperspectral cameras capable of sensing hundreds of wavelengths simultaneously in a streaming, real-time manner [Ungar et al. 2003]. Intelligent prioritization of the data for transmission to Earth is needed.

One solution to this problem is to employ onboard data analysis and classification methods that can automatically identify regions of interest. Many such detectors have been created to detect a variety of phenomena such as floods [Ip et al. 2006], volcanic eruptions [Davies et al. 2003], and dust devils imaged by Mars rovers [Castaño et al. 2008]. Beyond hand-crafted detectors, machine learning methods such as support vector machines (SVMs) have also been used onboard EO-1 to classify individual pixels into categories such as snow, water, ice, or land [Castaño et al. 2006]. SVMs are desirable for their compressibility into a simple dot product once trained in the linear case, maximizing computational efficiency in tightly constrained on-board environments. EO-1, for example, employs a 12 MHz MongOOSE M5 processor with 4 MIPS and 32 MB memory allocated to on-board science processing. The Autonomous Sciencecraft Experiment [Chien et al. 2005] is a highly visible, highly successful example of this technology that has demonstrably saved millions of dollars in operational costs and thereby enabled the Earth Observing One (EO-1) spacecraft to continue operating past its nominal end of mission date. We seek the automated detection of surface sulfur deposits, which on Earth have in some instances been linked to microbial activities. Using Borup Fiord as a case study, we evaluated a variety of classification and feature selection methods using data collected by the Hyperion instrument on EO-1.

The primary contribution of this paper is a careful description of an important real-world application of machine learning. First, we describe a concrete example of machine learning applied to a domain in which the both the general class of problem (automated interpretation of remotely sensed data) and the specific task (sulfur deposit detection) are of scientific interest (Sections 2 and 3). Second, we provide motivation for why a data-intensive, machine learning solution is needed by first attacking the problem with existing spectrum matching (Section 4.1) and endmember analysis (Section 4.2) methods. Third, we show how the sulfur detection problem requires an automated feature selection solution (Section 5). Finally, we discuss the process of subclass discovery and data filtering which ended up providing the key leverage point to achieve good performance at this task (Section 6.3).

Very few of these aspects have been addressed in previous work. The SVM devel-

oped for snow/water/ice/land (SWIL) classification [Castaño et al. 2006] addressed an easier classification problem and good performance was achieved by the SVM (surpassing expert hand-crafted models). However, no automated feature selection was explored as the observing bands were selected by an expert, and no additional data filtering was required again due to the relative simplicity of the task.

Early work on the use of SVMs for the sulfur detection problem was conducted by Castano et al. using a single Borup Fiord Hyperion scene from 2006 [Castaño et al. 2008]. Ground truth was provided by in situ collection of field spectra and laboratory spectral analysis of returned samples [Gleeson et al. 2010]. A later study augmented that data set with six subsequent observations that were made in 2007, increasing the generalizability and reliability of the trained classifier [Mandrake et al. 2009]. Neither report addressed the fundamental motivation issue. Finally, Section 7.2 contributes a new discussion of the implications of this study specifically for the study of Europa.

The classifiers described in this work were designed to run onboard the spacecraft to enable fast detection not only for Earth orbiters but also for future missions to remote destinations such as Europa. In a bandwidth-constrained setting, a Europa orbiter would be able to collect more data than it could transmit back to Earth. An onboard classifier could adaptively prioritize the regions of highest interest, including any similar sulfur deposits which could potentially point to sites of endogenous activity. Positive detections can also be used to adaptively replan spacecraft operations, e.g., to target a higher-resolution instrument or conduct more frequent observations of a particular region.

An example usage case of this method would begin with sample imagery from Europa provided to a ground team for analysis. Experts would employ other methods to discover regions of interest of known or unknown composition as well as regions of disinterest. An SVM classifier would then be automatically generated on the ground as described in this paper, designed to detect the presence of these interesting spectral features. Finally, the classifier would be uploaded to the Europa orbiter to help it prioritize images in real time that contain features of interest for priority transmission back to Earth and perhaps re-target such areas at higher spatial and spectral resolutions.

2. CASE STUDY: BORUP FIORD PASS, ELLESMERE ISLAND

Set in the far north of the Canadian Arctic Archipelago (81°N, 81°W), Ellesmere Island abuts the northern edge of Greenland between the Lincoln Sea and Baffin Bay. Like other ice-covered terrain in the solar system, it does not appear at first glance to be particularly hospitable to life. This region was explored in the mid 1990s by Benoit Beauchamp, who collaborated with later investigations in 2006 with Stephen Grasby, Damhnait Gleeson, and Marie-Eve Caron. His initial investigation discovered yellow stains along a glacial edge. These were identified as supraglacial deposits associated with sulfur-rich springs flowing through and across a 200-meter thick glacier [Grasby et al. 2003b]. The telltale yellow stain is due to the presence of elemental sulfur, which dominates the composition of the deposits. The source of these minerals most likely relates to marine evaporite deposits of anhydrite (CaSO₄) in the region's bedrock. The various depositional structures and spring locations

varied greatly between field campaigns. As Borup Fiord Pass is only accessible for three months of the year, the activation cycles of the springs during the remaining year are not known nor are the conditions that spur their flows.

The springs and their associated deposits show evidence of elemental sulfur, gypsum, and hydrogen sulfide gas. These three compounds contain sulfur in three different oxidation states. The presence of a complex redox system in addition to the high abundance of elemental sulfur (which is not predicted to be thermodynamically stable in the wet, oxygen-rich surface environment) suggested that microbial mediation of the local geochemistry could be occurring. Exploration of the microbial ecology within the system has detected a rich microfauna including known sulfur cycling microorganisms, while cultivation of microorganisms from this site has demonstrated that members of the community have the capacity to biomineralize sulfur [Grasby et al. 2003a; Gleeson et al. 2011]. This arrangement of fissured permanent ice through which sulfur-rich fluids access the surface and precipitate mineral deposits is potentially analogous to the surface of Jupiter’s moon Europa. Europa too shows evidence of surface sulfur-bearing mineral deposits on otherwise solid ice along fractures [Carlson et al. 1999]. The reddish-tinged lenticulae (round-shaped regions indicative of endogenic activity) and long, stained, double-lined fissures suggest a subsurface ocean likely rich with dissolved minerals [Sotin et al. 2002]. Studying the microbes surviving within the Borup Fiord glacial spring system might therefore also shed light on potential habitats beneath Europa’s surface, which would also be subject to low temperatures and potentially sulfur-rich conditions. The ability to automatically detect mineral deposits related to endogenic activity from orbit around Europa could greatly assist any future exploration of cryospheric biohabitats.

3. DATA SOURCE: HYPERION ON EO-1

Hyperion is a hyperspectral imager onboard the Earth Observing One spacecraft (EO-1). At a 700-km orbital altitude, Hyperion resolves the Earth’s surface to 30 m resolution simultaneously in 220 distinct spectral bands that range from 0.38 to 2.5 μm (near ultraviolet to short-wave infrared) [Ungar et al. 2003]. The first hyperspectral sensor to operate from space, it captures 7.5 km by 100 km of surface area with each image. Designed to assist mining, geology, forestry, agriculture, and environmental management through the classification of surface type and features, it has successfully generated over 27 Tb. of raw data available as various products from the USGS. In this study, we used the L1R Hyperion product, a processed version of the data that reports surface reflectance (as opposed to radiance) to help mitigate such effects as sun angle and reflections.

Our goal was to train a classifier that could accurately detect individual pixels containing sulfur. Damnhait Gleeson labeled a collection of pixels from seven Hyperion images using the commercial ENVI toolkit, aerial photos, and the Pixel-Learn image labeling software developed by the Machine Learning and Instrument Autonomy Group at the Jet Propulsion Laboratory [Mazzoni et al. 2005]. Each Hyperion image is 256 pixels wide and of variable length (mean 3200 rows); see Figure 1. Table I lists the Hyperion image IDs used along with the number of pixels labeled as sulfur, ice, and rock (the “bright” and “dark” distinction for sulfur

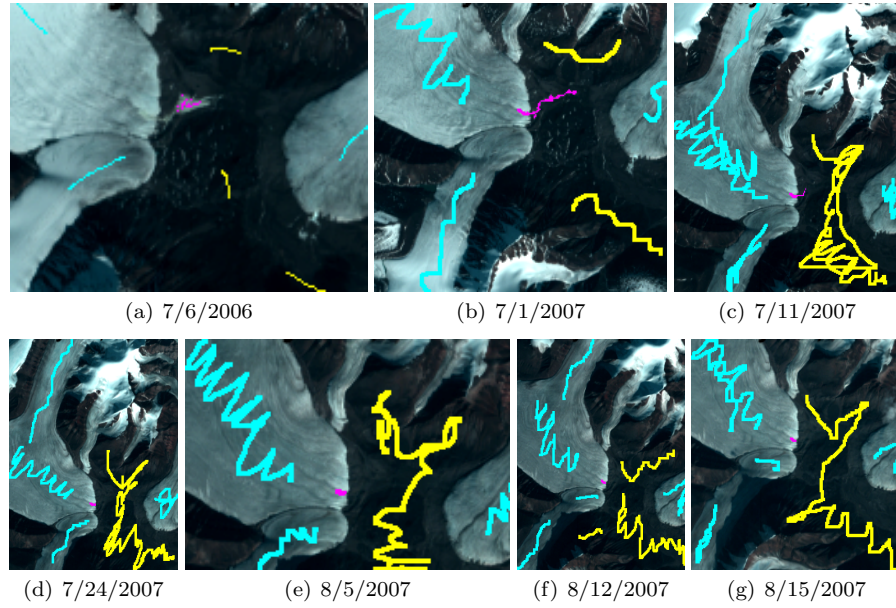


Fig. 1. Expert-provided labels for Hyperion images of Borup Fiord. Sulfur springs are at the right edge of the glacier. Key: violet = sulfur, cyan = ice, yellow = rock.

Table I. Number of sulfur, ice, and rock labels provided for seven Hyperion images of Borup Fiord.

Scene ID	Date	Sulfur Labels (Bright,Dark)	Ice Labels	Rock Labels
EO1H0570012006187110PY	7/6/06	18 (17,1)	75	58
EO1H0570012007182110KF	7/1/07	100 (26,74)	1322	579
EO1H0570012007192110KF	7/11/07	35 (22,13)	3016	2357
EO1H0570012007205110KF	7/24/07	25 (22,3)	2146	2067
EO1H0570012007217110KF	8/5/07	22 (22,0)	1643	1186
EO1H0570012007224110KF	8/12/07	19 (19,0)	2027	1712
EO1H0570012007227110PF	8/15/07	16 (16,0)	1800	1526
Total		235 (144,91)	12,029	9485

pixels is explained below). Sulfur is a rare occurrence, and the majority of each image remained unlabeled. The ice and rock labels served to provide non-sulfur examples for training the classifier. To evaluate false positive rates, we also used seven additional Hyperion images of similar size that targeted areas far from the sulfur-bearing glacial springs. These images were selected to cover a variety of polar regions with an environment similar to that of Borup Fiord, but without any known sulfur (see Table II).

An initial exploration of the labeled data shows that separating sulfur-bearing pixels from non-sulfur pixels is a challenging goal. Figure 2 plots all of the labeled pixels in terms of their reflectance for an arbitrary choice of two of Hyperion’s 220 wavelengths (similar results were observed for other wavelength pairs). The sulfur

Table II. Sulfur-free Hyperion images.

Scene ID	Date	Location
EO1H0692442004121110KZ	4/30/04	War Hunt Ice Shelf, Arctic
EO1H0630112006093110KF	4/3/06	Tuktoyaktuk, Arctic
EO1H0451162006327110PF	11/23/06	Ross Ice Shelf 1, Antarctica
EO1H2161062007085110KF	3/26/07	Larsen Ice Shelf, Antarctica
EO1H0201122008017110KF	1/17/08	Getz Ice Shelf, Antarctica
EO1H0270062008086110KF	3/26/08	Tyr 76, Arctic
EO1H0300052008101110KF	4/10/08	Thule AFB, Arctic

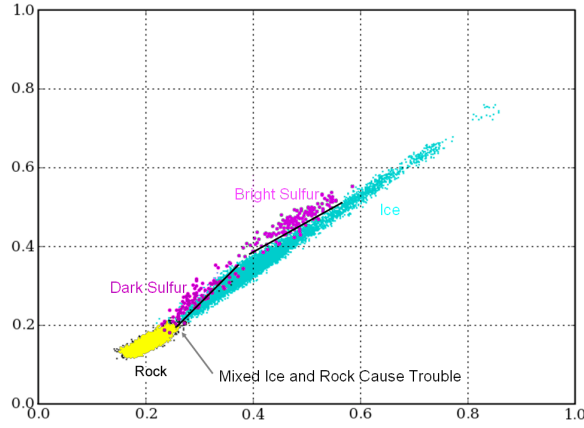


Fig. 2. Scatter plot of all 21,759 labeled pixels, colored by label type, using two of Hyperion’s 220 wavelengths for the x and y axes.

pixels are well mixed with the ice pixels, and there is some overlap with the rock pixels as well. Further, two sub-populations are evident in the pixels labeled as sulfur. Note that the “Dark Sulfur” points are especially easy to confuse with rock and similarly dark ice.

We used the k-means clustering algorithm [MacQueen 1967] to automatically refine the labeled sulfur pixels to account for these two subtypes. The two sub-populations were strongly distinguished by their mean reflectance values. We therefore refer to these types as “Bright Sulfur” and “Dark Sulfur” pixels and use their k-means labels to separate them in the labeled training data. We have confirmed empirically that this distinction yields better classification performance, and it also permits a more informed analysis of the results.

4. SULFUR DETECTION METHODS

4.1 Sulfur Spectrum Matching

As a starting point, we first compared the pixel spectra observed by Hyperion to the standard USGS lab spectrum for sulfur [Clark et al. 2007]. This spectrum, restricted to the range of wavelengths that Hyperion uses, is shown in Figure 3. It is a mostly flat and bright spectrum across the near-infrared with a steep edge in the visible range.

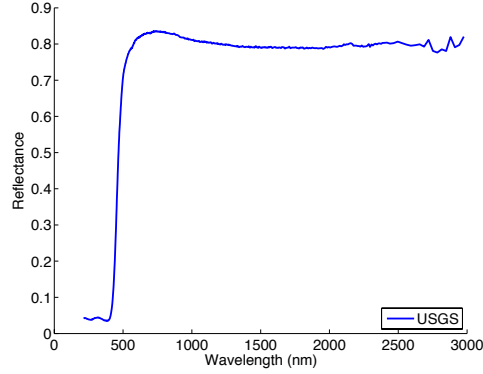


Fig. 3. USGS sulfur lab spectrum, restricted to Hyperion's observing wavelengths

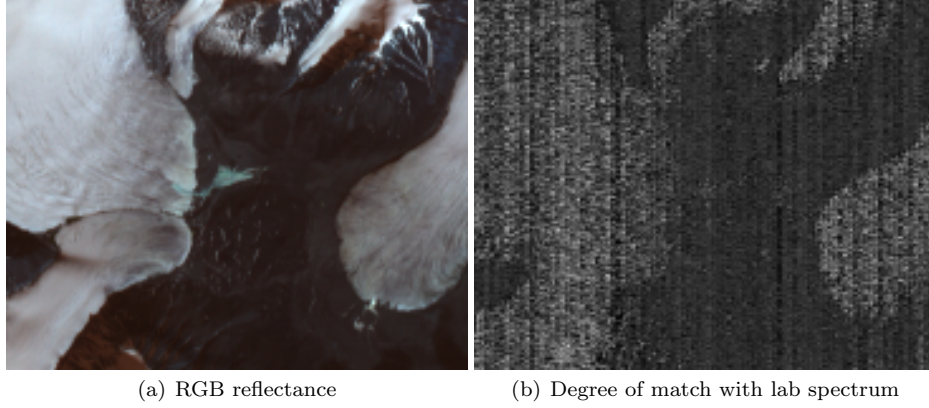


Fig. 4. Sulfur lab spectrum match results for EO1H0570012006187110PY

We computed the degree of sulfur spectrum match for each pixel i as

$$s(i) = h_i \cdot S$$

where h_i is the vector of Hyperion reflectances observed for pixel i and S is the vector of reflectance values for the sulfur lab spectrum. Figure 4(a) shows an RGB representation of the Borup Fiord area from image EO1H0570012006187110PY. The sulfur spring appears near the center of the image as an outflow from the glacier tongue on the left. Figure 4(b) shows the sulfur spectrum match results for each pixel, in which brightness correlates with a higher degree of match. The highest matches tend to be the icy areas, since ice also has a relatively flat, bright spectrum. The sulfur deposit is not recovered. We note significant instrument noise, with vertical striping due to variabilities across detector elements and high levels of background noise due to single-pixel artifacts at high wavelengths [Pearlman et al. 2003]. Additionally, given Hyperion's 30-m spatial resolution it is likely that pixels with sulfur also contain other elements such as ice and rock that preclude a direct

matching with the (pure) lab spectrum.

This result suggests two directions for improvement: first, adopting a data-driven approach that adapts sulfur detection to the actual spectra observed, given instrument properties and subpixel mixing (addressed next); and second, carefully selecting the most informative and reliable bands for conducting the analysis (addressed in Section 5).

4.2 Endmember Analysis

Our next step was to employ an unsupervised data analysis technique to investigate how separable the sulfur pixels were from the surrounding ice and rock features. Linear *endmember detection* assumes that spectra are linear combinations of pure materials, and that these minerals combine according to nonnegative mixing coefficients to produce all of the observed spectra in the scene [Keshava and Mustard 2002]. Examples of endmember detection algorithms include Sequential Maximum Angle Convex Cone Analysis (SMACC) [Gruninger et al. 2004], Vertex Component Analysis [Nascimento and Dias 2005], and N-FINDR [Winter 1999]. They compute geometric or statistical properties of the data to determine the endmember spectra without any strong assumptions about the specific classes of materials that are present. They thereby provide an objective test of the prominence of the sulfur signal with respect to the background.

Endmember detection is sensitive to spectral artifacts, so its application to orbital remote sensing requires careful attention to mitigate noise. We employ a superpixel endmember detection approach that incorporates a segmented image representation to help compensate for pixel noise [Thompson et al. 2010]. We begin with a conservative oversegmentation of the scene into contiguous, homogeneous pixel groups. These groups are created by agglomerating pixels together according to a spectral similarity measure. Here a simple Euclidean distance measure informs a graph partitioning segmentation algorithm [Felzenszwalb and Huttenlocher 2004]. The result of the operation is a set of regions that are spatially contiguous and spectrally homogeneous; the new superpixel representation allows the mean spectrum of each superpixel to stand in for the entire region. This provides noise reduction benefits while generating regions of interest for later processing.

We applied this endmember analysis to a superpixel representation of the same Hyperion image (EO1H0570012006187110PY) to determine whether sulfur would emerge as a distinct endmember. We used the SMACC strategy [Gruninger et al. 2004], which sequentially grows a set of endmembers to maximize the volume of the implied convex cone. This provides a rank ordering of endmembers in terms of their extremity *vis a vis* the remaining data. Because the sulfur signal was known to occupy visible bands, we focused on this region of the spectrum. Our spectral analysis used only the Hyperion bands in the range 426 to 626 nm. This excluded the lowest frequency channels which generally had more instrument artifacts. Pixels from the sulfur region appeared in the 15th endmember. Figure 5(a) shows a spatial map of where the pixels from the 15th endmember were located (green). Figure 5(b) shows the average spectrum for those pixels, with the USGS lab spectrum for sulfur as a reference. The characteristics of the sulfur spectral signal make it a prominent signal in this data set. However, it also deviates significantly from the pure lab spectrum for sulfur, which is consistent with the lab spectrum matching results

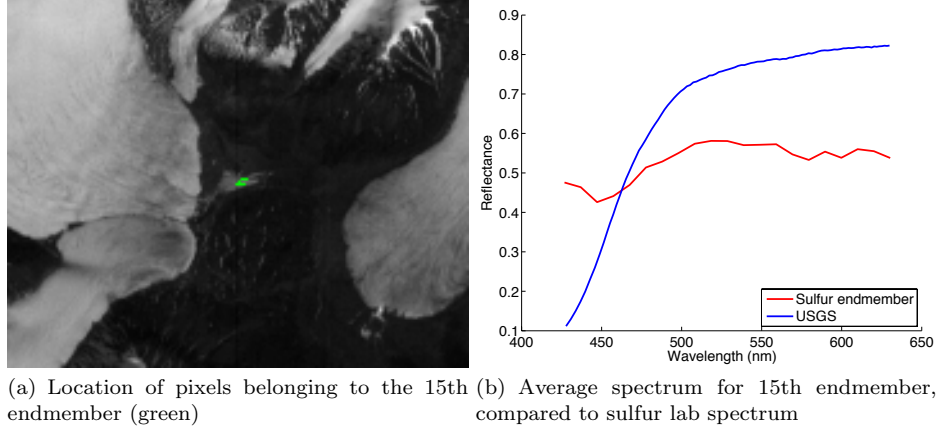


Fig. 5. Endmember analysis results for EO1H0570012006187110PY

presented earlier. The sulfur signal is filtered through the observing instrument’s properties, the environmental conditions, and subpixel mixing; these factors are all ignored by the endmember detection process.

We conclude that the sulfur signal is strong, but too scene-specific to compare directly with the lab spectrum. Thus we are naturally guided to a supervised classification solution, which is described next.

4.3 Supervised Training of a Sulfur Classifier

Supervised classification methods make use of a small collection of expert-labeled data to train a generalized model that can make similar classification decisions on new data. In using the labeled data, rather than lab spectra, we have access to examples of sulfur (and non-sulfur) that is tailored to the instrument and observing conditions.

The classifier we used for this study was the support vector machine (SVM) [Cortes and Vapnik 1995]. SVMs generally provide high performance across a wide range of problems and have been used previously for sulfur classification with Hyperion, albeit with a very small data set (only 18 labeled sulfur pixels) [Castaño et al. 2008]. Another relevant benefit (for use onboard a spacecraft) the SVMs provide is their great computational efficiency, especially with linear kernel implementations. An SVM seeks to construct a hyperplane that separates two classes of previously labeled data. A linear kernel SVM is defined by $n+1$ parameters: a weight (w_i) for each of n training examples (x_i) and a single scalar bias term (b). These weights and bias are determined during the training process by solving a quadratic programming problem with one degree of freedom: the regularization parameter C which determines how strongly to penalize training errors. Classification of a new example x as positive or negative is achieved by:

$$F(x) = \text{sign} \left(\sum_{i=1}^n w_i (x_i \cdot x - b) \right).$$

More expressive solutions can be obtained by replacing the linear dot product with a kernel function $K(x_i, x)$ that implicitly maps each point into a new (possibly infinite dimensional) feature space in which linear separability is achievable. For this study, we examined both a linear kernel (simple dot product) and a Gaussian kernel parameterized by γ , the Gaussian width:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{\gamma}\right). \quad (1)$$

The Gaussian kernel generally provides higher accuracy and greater representation power, but it requires significantly more computational effort. It serves in this study as an example of what might be achievable given more spacecraft resources than are generally available on EO-1.

Before presenting classification results using an SVM trained from the expert-provided labeled data, we will next address the second important advance required for this work: an intelligent selection of which of Hyperion’s 220 observational bands to use when seeking to classify sulfur pixels.

5. FEATURE SELECTION

Although Hyperion generates 220 bands of hyperspectral data, only 12 (selectable) bands can be passed to the onboard processor for analysis due to architectural limitations. Therefore, any onboard classifier faces the dual challenges of choosing which bands are most relevant for its classification goals and then classifying the data as accurately as possible. The instrument exhibits a sharp decrease in data quality at higher wavelengths (band numbers) [Pearlman et al. 2003], so the most reliable bands are likely to be the lower ones.

We employed the following methods for feature selection:

- (1) Domain expert manual selection based on knowledge and experience.
- (2) Greedy Forward Feature Selection [Guyon et al. 2002]: Starting with the empty set, incrementally and greedily add the feature that yields the highest performance on a held-out validation set, until the desired number of features have been selected. While simple to implement, this strategy cannot accommodate dependencies between features (e.g., it scores each feature in isolation and is therefore blind to the advantages of certain feature pairs or groups).
- (3) Recursive Feature Elimination (RFE) [Guyon et al. 2002]: Starting with the full set of available features, recursively and greedily remove the feature that results in the smallest decrease in performance. Stop when the desired number of features remain. This approach can retain groups of features that are more useful together than they appear to be individually.
- (4) RFE based only on the 2006 data [Castaño et al. 2008]: Same as (2) above, but using only the single source image from 2006. This is all that was previously available as labeled data for this region and is included for comparison.

Each of these methods was applied to the (labeled) Hyperion training data, with the goal of selecting just 12 bands. Table III shows the bands (in terms of their center observing wavelengths) that were selected by each method, sorted by wavelength. The first four bands were the same for all methods. The final row in the

Table III. Hyperion band selection results (in nm). Items in bold were unique to that feature selection method.

	Expert Pick	Greedy Forward	RFE 2006-2007	RFE 2006 only
Wavelengths chosen	426	426	426	426
	436	436	436	436
	446	446	446	446
	456	456	456	456
	466	556	466	506
	486	566	536	516
	506	576	546	526
	526	586	556	536
	546	596	566	546
	566	1448	576	556
	586	1791	586	576
	626	2013	596	994
Number matching expert	12	6	8	7

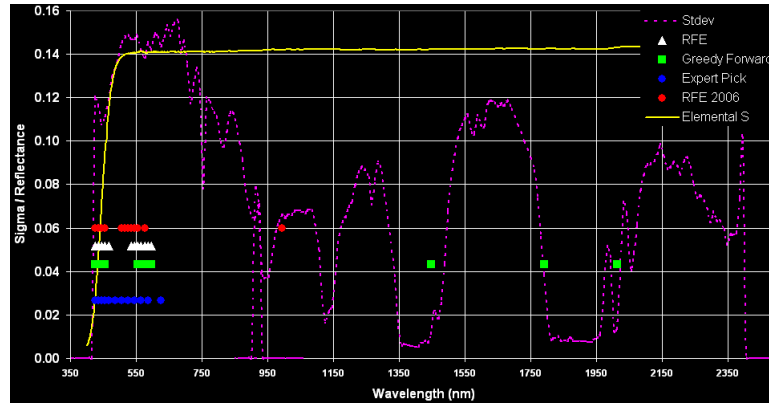


Fig. 6. Hyperion band selection results, shown with sulfur lab spectrum (normalized so 100% reflectance is at the top of the graph) and the amount of variance per band (as a rough proxy for instrument sensitivity). Bands chosen by each feature selection method are marked with an arbitrary y value.

table indicates how many of the bands selected were also chosen by the expert. RFE using all seven training scenes showed the strongest agreement with the expert (8 of 12 bands). Bands that were unique to a single selection method are shown in bold. The greedy forward method is distinguished in its use of longer wavelengths (1448, 1791, and 2013 nm).

Figure 6 shows the band choices visually, using a different marker for each feature selection method. The figure also shows the sulfur lab spectrum across the same wavelength range, and the standard deviation observed across the data set at each wavelength. This value serves as a rough proxy for the sensitivity of the Hyperion instrument [Pearlman et al. 2003] and is largely influenced by atmospheric absorption properties. Most of the bands chosen also have high standard deviation values,

except for the three highest bands chosen by the Greedy Forward method, all of which have much lower sensitivity. This combined with the high computational cost for the Greedy Forward strategy led us to focus on the other three options for the remainder of the paper.

Sequential band selection shows an interesting parallel with many manually derived spectral classification rules. Many mineralogical maps exploit specific scalar operators that reveal the presence of specific absorption features. For example, the CRISM spectral summary products include an ISLOPE index [Pelkey et al. 2007] that is sensitive to ferric coatings on dark rock in Mars CRISM images. It is a scalar value generated by measuring the spectral slope between 1815 and 2530 nm. The algorithm estimates the reflectance values R_{1815} and R_{2530} at each point and computes the numerical slope between them. Other spectral indices compute band depths relative to a continuum value.

The RFE sequential selection algorithm “discovers” a similar strategy, choosing clusters of bands that provide reliable estimate reflectances at 436 and 536 nm. In other words, it naturally brackets the sulfur’s spectral slope feature. Forcing the band selection process to select fewer than 12 bands tends to alternately add bands to either side of the dynamic region, refining the effective estimate of the sulfur feature’s depth. This oscillation continues until it has exhausted all spectral bands near 436 and 536nm. This policy is optimal for reducing the uncertainty over the distribution of absorption depths that could have given rise to the observed feature vector. We take our estimate of the two reflectances to be Gaussian-distributed random variables $\mathcal{N}(\mu_a, \sigma_a^2)$ at 436nm and $\mathcal{N}(\mu_b, \sigma_b^2)$ at 536nm. This is reasonable for white instrument noise. After taking i.i.d. measurements X_a from the 436 nm region and measurements X_b from the 536 nm region, the posterior variance of the reflectance estimates reduces according to the “root-n” rule:

$$P(R_{436}|X_a) = \mathcal{N}(\mu_a, \frac{\sigma_a^2}{|X_a|}), \quad P(R_{536}|X_b) = \mathcal{N}(\mu_b, \frac{\sigma_b^2}{|X_b|}).$$

The depth of the sulfur feature is the difference of the two reference values:

$$f(436, 536) \sim \mathcal{N}(\mu_b, \sigma_b^2) - \mathcal{N}(\mu_a, \sigma_a^2)$$

This linear combination of Gaussians is also Gaussian-distributed with variance $\sigma^2 = \sigma_{436}^2 + \sigma_{536}^2$. We quantify its uncertainty with the (differential) Shannon entropy:

$$h(S|X) = \frac{1}{2} \ln \left(2\pi e \frac{\sigma_a^2}{|x_a|} \right) + \frac{1}{2} \ln \left(2\pi e \frac{\sigma_b^2}{|x_b|} \right)$$

For some constant C , this is equal to the expression:

$$h(S|X) = C - \frac{\ln |x_a| - \ln |x_b|}{2}$$

Additional measurements from either side yield diminishing returns for the depth estimate due to the natural logarithm. For maximum reduction in entropy it is always desirable to allocate measurements to the reflectance term with the least number of previous observations. By alternating between low and high bands, the greedy RFE band selection seeks an “elemental sulfur index” value representing the

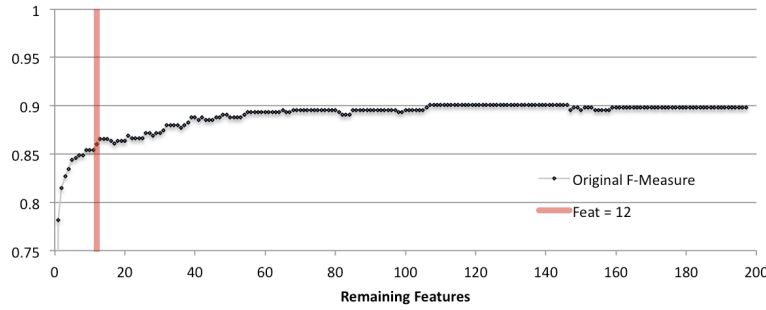


Fig. 7. RFE band selection results for a variety of number of features selected. Performance declines as fewer features remain. The red bar marks performance with 12 features selected.

best information gain with respect to the depth of the sulfur absorption feature. It discovers this strategy using only noisy labeled examples of the data, but no other prior information about the spectral characteristics of sulfur.

The number of bands selected (12) is dictated by the architecture of the EO-1 spacecraft in this application. It is reasonable to wonder how much performance is lost due to the discarded bands. Figure 7 shows f-measure performance as a function of the number of bands remaining when selected by the RFE method. Recall that RFE starts with all features and then iteratively removes the least useful one. We observe that there is effectively no impact on performance until the number of bands drops below 60. At the selected performance point of 12 bands, the f-measure achieved is 0.86, only 0.04 below the asymptotic performance of 0.90 achieved with the full set of bands. We therefore conclude that there is a little room for improvement if more bands could be used, but the impact to performance is not dramatic in restricting the analysis to only 12 bands. Using fewer than five bands, however, would dramatically impact performance.

6. CLASSIFICATION RESULTS

For each of the four 12-band feature sets chosen by the methods above, we performed a cross-validation study by training an SVM on the pixels from all but one of the labeled scenes and computing test performance on the pixels from the remaining scene, then averaging results across the test scenes.

6.1 SVM Performance Metrics

To evaluate performance, we are interested both in accuracy and in the number of false positives. False detections, if used to trigger followup observations of a target area, can be very expensive, so it is important that they be kept low.

For accuracy, we report the f-measure, which is the harmonic mean of recall (R) and precision (P) [Van Rijsbergen 1979]. Here, L indicates labeled pixels and D indicates detected pixels (the output of the classifier). L and D are subscripted by their type of label (class): I = ice, R = rock, B = bright sulfur, and D = dark

sulfur.

$$\begin{aligned}
S_{correct} &= |L_B \cap (D_B \cup D_D)| \\
S_{missed} &= |L_B \cap (D_I \cup D_R)| \\
S_{false} &= |D_B \cap (L_I \cup L_R)| \\
P &= \frac{S_{correct}}{S_{correct} + S_{false}} \\
R &= \frac{S_{correct}}{S_{correct} + S_{missed}} \\
F &= \frac{2PR}{P + R}
\end{aligned}$$

These metrics focus on the accurate detection of the bright sulfur class, which is the one most likely to confidently contain surface sulfur. $S_{correct}$ is the number of correct detections of bright sulfur (whether detected as bright or dark sulfur by the classifier). S_{missed} is the number of missed detections of labeled bright sulfur (identified as anything but sulfur), and S_{false} is the number of incorrect bright sulfur identifications (ice or rock identified as bright sulfur).

We also compute two other statistics, both addressing false positives identified outside the original labeled pixels. Any unlabeled pixels in the seven training images that are detected as sulfur are considered “likely false positives” and referred to as *Likely FPOS* in our graphs. In the absence of a label or ground truth, we cannot know with certainty that there is no sulfur present, but it is extremely unlikely. We also compute the number of *Sulfur-Free FPOS*, which is the number of pixels in the seven new scenes taken very far from the known sulfur source (images of clouds, opposite hemisphere ice sheets, frozen ocean, etc.). Any putative sulfur detected in the Sulfur-Free images can be confidently considered an error.

6.2 Initial SVM Results

We explored three different formulations of the sulfur detection problem as a classification task:

- Two classes: sulfur (bright and dark) vs. non-sulfur (ice and rock)
- Three classes: sulfur (bright and dark), ice, rock
- Four classes: bright sulfur, dark sulfur, ice, and rock

For each formulation, we trained a multi-class SVM using a one-vs.-all ensemble of pairwise classifiers. In each case, we used a linear kernel to minimize computational cost, in light of the desired operation of the classifier onboard the EO-1 spacecraft. Later we examine the use of a Gaussian kernel.

The linear SVM has only a single hyperparameter, C , to tune. C trades off penalties for errors on the training data against model complexity; small values permit better generalization while large values more accurately fit the training data. We tested $C \in \{10^{-1}, \dots, 10^5\}$ using 36 values ranging logarithmically.

Figure 8 shows SVM performance results using 2, 3, or 4 classes to define the problem (top, middle, and bottom rows respectively). Each plot shows results for three different band selection strategies (expert pick, RFE old (from 2006, see

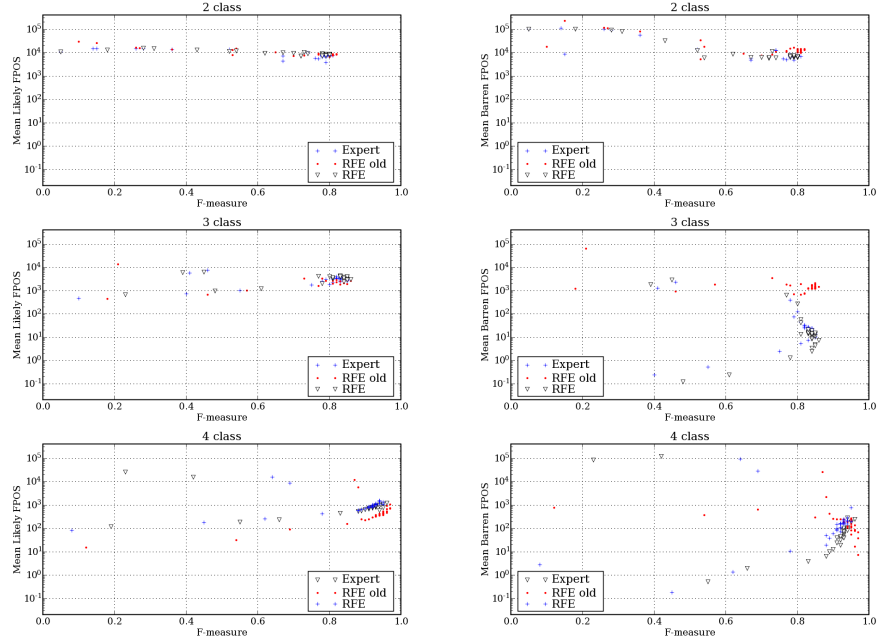


Fig. 8. SVM performance for 2, 3, and 4 class formulations, with different band selections, using a linear SVM. The x axis is f-measure ranging linearly from 0.0 to 1.0; the y axis shows the number of likely false positives from the same test image (left) or from distant sulfur-free images (right). Note that the y axis uses a logarithmic scale.

above), and RFE), for each of the C values tested. The left column plots Likely FPOS against f-measure, while the right column shows Sulfur-Free FPOS against f-measure. The optimal result would be lie at the bottom-right of each graph, with zero Likely / Sulfur-Free FPOS (Y) and an f-measure of one (X). Note that the y axis uses a logarithmic scale to report the number of false positives.

The linear kernel 2-class (sulfur vs. non-sulfur) SVM shows poor performance. While some C values yielded an impressive f-measure of up to 0.80-0.82 for all three band selection strategies, these classifiers also produced more than 104 Sulfur-Free FPOS. Similarly poor generalization occurred in the Likely FPOS of the unlabeled examples from the source image. Thus, though the labeled data was fit, it failed to generalize to more remote examples.

The linear kernel 3-class (ice, rock, and sulfur) SVM achieved an improvement for all three band selections, particularly for the Expert and RFE selected bands. The number of Sulfur-Free FPOS dropped to a mean of ~ 40 FPOS per Sulfur-Free image. The number of Likely FPOS remained high, however, with a mean of ~ 3400 per image. The mean and maximum f-measure also increased with a new top achievement of 0.86.

The linear kernel 4-class (ice, rock, bright sulfur, and dark sulfur) SVM yielded improvements in f-measure and mixed results with false positives. The mean f-measure increased to a mean of ~ 0.9 and a maximum of 0.96. The number of

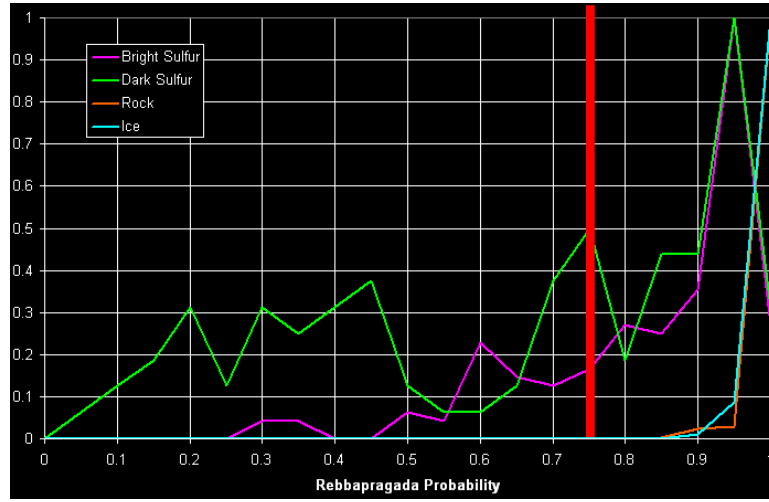


Fig. 9. Histogram of PWEM estimates of label correctness, broken down by labeled class. Everything left of the red line was filtered as being too unreliable.

Sulfur-Free FPOS increased to a mean of ~ 70 pixels per image, but the number of Likely FPOS was reduced to a mean of ~ 800 false positives per image. Most of Likely FPOS improvement comes from the decision to define recall and precision in terms of the system’s ability to detect only the bright sulfur.

6.3 Filtering the Labeled Data

Any evaluation is only as good as the reference that is used. As we worked with this data, it became clear that some of the labeled pixels were less reliable than others. Manual pixel labeling is a challenging task even for experts. The patterns of labeled pixels in Figure 1 suggest the strategy by which they were labeled, which included broad sweeps of a coarse “paint” tool across seemingly homogeneous areas for the ice and rock classes. The sulfur pixels were labeled more cautiously by zooming in and marking individual pixels. In all cases, subpixel mixing and unclear boundaries can render these labels noisy.

We therefore conducted an experiment to determine whether the noisy labels could be automatically identified and filtered to improve the quality of the labeled data. We employed Pair-Wise Expectation Maximization (PWEM), a label confidence evaluation method based on averaging across multiple EM clusterings of the labeled data [Rebbapragada and Brodley 2007; Rebbapragada et al. 2009]. In its general formulation, PWEM suggests that improved results may be obtained by permitting a reweighting of input labels based on these results. In this study, we simply filtered out any labeled example with an insufficiently high (0.75) probability of being labeled correctly.

Figure 9 shows the distribution of label correctness probabilities, broken down by pixel type (label). Ice and rock labels proved to be the most reliable (almost all near 1.0). When using the original three classes (bright and dark sulfur grouped together; results not shown in the figure), none of the “sulfur” pixels had a confidence that

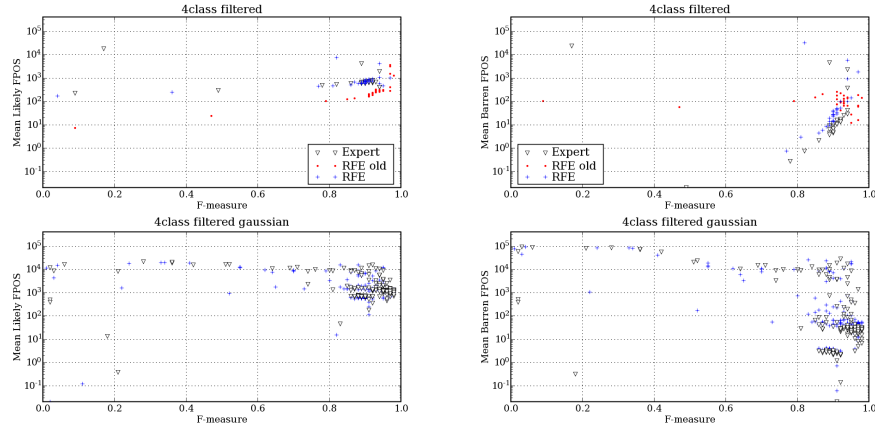


Fig. 10. SVM performance for the 4-class formulation, using different band selections and filtering (via PWEM) labeled items with a confidence of less than 0.75. The SVM kernel used was linear (top row) or Gaussian (bottom row). Axes are the same as in Figure 8.

exceeded 0.75. Separating these pixels into bright and dark sulfur provided a stark benefit: 104 of the 144 bright sulfur labels and 39 of the 91 dark sulfur labels had a confidence at or above 0.75. We then discarded labels with confidence below 0.75, a threshold chosen so as to provide a reasonably confident resulting labeled set without decreasing its size too much. The labeled pixels that were removed were largely located near the boundary between ice and rock areas.

Filtering out the likely mislabeled (noisy) items dramatically reduced the number of false positives. Revised results achieved using PWEM to estimate label confidence and then filter the labels are shown in Figure 10. The mean f-measure was unchanged (0.90) while the number of Likely FPOS dropped to ~ 600 per image. Sulfur-Free FPOS dropped even further to a mean of 24 pixels per image. While further filtration of the labels might in principle yield further improvements, we had already discarded 40% of our very limited positive examples for the gains shown here, and discarding more would risk overfitting the very small labeled data set.

6.4 Gaussian SVM Results

The final part of this study reports on the use of a Gaussian kernel, which provides increased representational power (with respect to a linear kernel) at the cost of additional computational effort. A second hyper-parameter, γ , must be specified; this parameter indicates the “width” of the exponential expression in Equation 1. In practical terms, it is not possible for the EO-1 spacecraft to make use of an SVM with a Gaussian kernel given processor limitations. However, looking ahead to future missions and technology, we studied this type of classifier in order to understand what benefits it may provide.

Our experiments with the Gaussian kernel used the same C values but also varied the Gaussian width parameter $\gamma \in \{10^{-2}, \dots, 10^3\}$ by factors of ten. These ranges spanned the empirically observed dynamic region between numeric instability and a failure of the SVM to generalize beyond the training data.

Figure 10 (bottom) shows the result of employing a Gaussian kernel on the PWEM-filtered 4-class data. Results are shown only for the Expert and RFE band selections. The plots encompass a larger number of results due to the additional sweep over values for γ . The Gaussian SVMs achieved a mean f-measure of 0.93 and maximum f-measure of 0.98, exceeding the best performance of the linear kernel SVM, without sacrificing performance in terms of the Likely and Sulfur-Free FPOS.

7. DISCUSSION AND CONCLUSIONS

The experimental results show that a classifier can indeed be trained to detect sulfur-bearing pixels in data collected by the Hyperion instrument onboard EO-1, while accommodating the particular computational challenges and constraints imposed by the onboard environment. The best results were achieved by 1) using Recursive Feature Elimination (RFE) to select the 12 bands to use, 2) modeling the problem using four classes by decomposing “sulfur” into “bright sulfur” and “dark sulfur” populations, and 3) employing PWEM to filter out likely mislabeled items from the training set.

Automated feature selection is effective on this problem. RFE tended to select bands similar, but not identical, to those chosen manually by an expert, and RFE’s bands often yielded higher f-measure performance as well as a smaller number of false positive detections.

Identifying the distinct subpopulations within the sulfur class not only yielded improved generalization performance, but it also provides an example of how the result of a machine learning investigation can feed back into the underlying science goals of an investigation. Empirically, the machine learning results indicate that the distinction between sulfur on ice and sulfur on rock is important for creating a reliable model of sulfur in this environment. The use of PWEM to estimate the confidence of each label could provide additional benefits. We employed this information only to filter out unreliable labels, but it could also be used as feedback to the expert who originally created the labels. While reviewing each individual labeled pixel would be exceedingly tedious, it is reasonable to simply review a small number of pixels that were marked as questionable. Then, rather than filtering them out, we could make use of those that the expert considers truly reliable, even if unusual or anomalous with respect to the rest of the training data. The value of PWEM would then be in guiding the “second look” efforts of the expert.

The best sulfur classification performance we observed was an f-measure of 0.9 with ~ 600 likely false positives per training image and just 24 false positives per sulfur-free image. These results were achieved with a highly efficient linear kernel SVM, which is a classifier that can be run onboard EO-1 in approximately 50 minutes. We also found that a Gaussian SVM could achieve higher performance (f-measure 0.98), but at a computational cost that is unrealistic for use on EO-1 (24 hours to complete). This classifier could however be employed on a future mission with a more capable onboard processor.

7.1 Understanding the False Positives

In this application, false positive detections are very undesirable since detections may be used to schedule costly follow-up observations. We therefore investigated

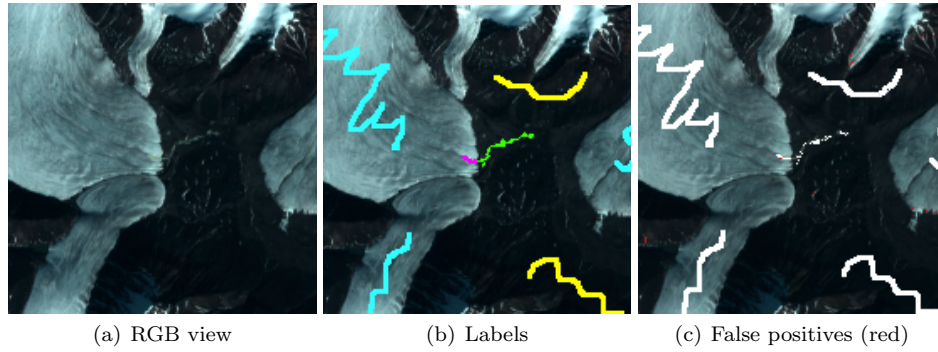


Fig. 11. Example of Likely False Positive detections for image EO1H0570012007182110KF. The figure shows an RGB view (left), annotated with labels including the separation of sulfur into bright (magenta pixels) and dark (green) sulfur (middle), and the false positives as red pixels (right).

the false positives more closely to help understand their prevalence. Figure 11 shows examples of the false detections obtained from a highly performing classifier. (SVM with Gaussian kernel, $C = 0.25$, $\gamma = 1000$, RFE band selection). The test accuracy achieved by this classifier was an f-measure of 0.83, 0 sulfur-free FPOS, and ~ 50 Likely FPOS per image. The left panel of the figure shows the original image in RGB. The middle panel shows the data labels for ice (cyan), rock (yellow), bright sulfur on ice (magenta), and dark sulfur on rock (green). The right panel marks in white all labeled pixels that remain after applying PWEM confidence estimation and filtration, and in red the false detections that comprise the Likely FPOS for this image. These tend to occur in small icy valleys and at glacier edges, which are environmentally similar to the true sulfur locations but do not contain sulfur itself. As noted earlier, the spectra for sulfur and for ice are quite similar.

This analysis highlights a potential weakness in the methodology involved in obtaining labeled data from an expert user. If the labels that are provided include unambiguous examples (such as our ice and rock labels) as well as more challenging examples of a minority class that coexists with ambiguous examples of the other classes, then the trained classifier may learn to highlight “mixed” areas as being the ones of interest, not those that simply contain the minority class.

7.2 Implications for Europa

The results of this terrestrial case study indicate that laboratory spectra of materials collected under ideal conditions can deviate significantly from the spectra of materials as they are observed in the field. Incorporating the experience gained in developing data-derived solutions for the identification of materials of interest to Europa will be particularly valuable, especially given the topographically variable conditions and low signal to noise data sets utilized for the current study. While very smooth at the global scale, the topography of Europa at higher resolution is observed to be highly variable, with successive generations of intersecting linear geologic features recorded across the icy surface in the form of larger ridges and furrows, among countless smaller cracks and lineations. The spatial resolution of

current datasets average 30 km per pixel, and while future missions will improve this resolution, pixels are likely to contain multiple materials, complicating spectral signatures. Onboard algorithms that can carry out automated classification could be utilized to intelligently select sites for higher resolution or repeat data collection. Definitive identification of sites of sulfur-rich materials on the surface of Europa related to endogenous activity could be considered as candidates for future landing sites.

ACKNOWLEDGMENTS

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. © 2011. Special thanks to the JPL Supercomputing and Visualization Facility for their support. Expert label generation and analysis was carried out by Damhnait Gleeson under the NASA Planetary Geology and Geophysics Program grant NNX07AR28G to Dr. Robert T. Pappalardo.

REFERENCES

- CARLSON, R. W., JOHNSON, R. E., AND ANDERSON, M. S. 1999. Sulfuric acid on Europa and the radiolytic sulfur cycle. *Science* 286, 5437, 97–99.
- CASTAÑO, A., FUKUNAGA, A., BIESIADECKI, J., NEAKRASE, L., WHELLEY, P., GREELEY, R., LEMON, M., CASTAÑO, R., AND CHIEN, S. October 2008. Automatic detection of dust devils and clouds at mars. *Machine Vision and Applications* 19, 5-6, 467–482.
- CASTAÑO, R., MAZZONI, D., TANG, N., GREELEY, R., DOGGETT, T., CICHY, B., CHIEN, S., AND DAVIES, A. 2006. Onboard classifiers for science event detection on a remote sensing spacecraft. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 845–851.
- CASTAÑO, R., WAGSTAFF, K., GLEESON, D., CHIEN, S., TRAN, D., SCHARENBRUCH, L., MODHADAM, B., TANG, B., BUE, B., DOGGETT, T., MANDL, D., AND FRYE, S. 2008. Onboard detection of active Canadian sulfur springs: A Europa analogue. In *Proceedings of the International Symposium on Artificial Intelligence, Robotics, and Automation in Space (iSAIRAS)*.
- CHEUNG, K.-M., DIVSALAR, D., DOLINAR, S., ONYSZCHUK, I., POLLARA, F., AND SWANSON, L. 1993. Changing the coding system on a spacecraft in flight. In *IEEE International Symposium on Information Theory*. 381.
- CHIEN, S., SHERWOOD, R., TRAN, D., CICHY, B., RABIDEAU, G., CASTANO, R., DAVIS, A., AND BOYER, D. 2005. Using autonomy flight software to improve science return on Earth Observing One. *Journal of Aerospace Computing, Information, and Communication* 2, 196–216.
- CLARK, R. N., SWAYZE, G. A., WISE, R., LIVO, E., HOEFEN, T., KOKALY, R., AND SUTLEY, S. J. 2007. USGS digital spectral library splib06a. U.S. Geological Survey, Digital Data Series 231.
- CORTES, C. AND VAPNIK, V. 1995. Support-vector network. *Machine Learning* 20, 273–297.
- DAVIES, A. G., BAKER, V., CASTAÑO, R., CHIEN, S., CICHY, B., DOGGETT, T., DOHM, J., GREELEY, R., RABIDEAU, G., SHERWOOD, R., WILLIAMS, K., AND TEAM, A. P. 2003. Autonomous volcanic activity detection with ASE on EO-1 Hyperion: Applications for planetary missions. *Bulletin of the American Astronomical Society* 35.
- FELZENSZWALB, P. F. AND HUTTENLOCHER, D. P. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59, 2 (September), 167–181.
- GLEESON, D., GRASBY, S., WILLIAMSON, C., SPEAR, J., PAPPALARDO, R., AND TEMPLETON, A. 2011. Low temperature S0 biomineralization at a supraglacial spring system in the Canadian High Arctic. *Geobiology* 9, 360–375.
- GLEESON, D., PAPPALARDO, R. T., GRASBY, S. E., ANDERSON, M. S., BEAUCHAMP, B., CASTAÑO, R., CHIEN, S., DOGGETT, T., MANDRAKE, L., AND WAGSTAFF, K. 2010. Characterization of
- Journal of the ACM, Vol. V, No. N, Month 20YY.

- a sulfur-rich Arctic spring site and field analog to Europa using hyperspectral data. *Remote Sensing of Environment* 114, 1297–1311.
- GOLDSTEIN, R. M., ENGELHARDT, H., KAMB, B., AND FROLICH, R. M. 1993. Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* 262, 5139, 1525–1530.
- GRASBY, S. E., ALLEN, C. C., LONGAZO, T. G., LISLE, J. T., GRIFFIN, D. W., AND BEAUCHAMP, B. 2003a. Biogeochemical sulphur cycle in an extreme environment—life beneath a high arctic glacier, Nunavut, Canada. *Journal of Geochemical Exploration* 78–79, 71–74.
- GRASBY, S. E., ALLEN, C. C., LONGAZO, T. G., LISLE, J. T., GRIFFIN, D. W., AND BEAUCHAMP, B. 2003b. Supraglacial sulfur springs and associated biological activity in the Canadian high Arctic: signs of life beneath the ice. *Astrobiology* 3, 3, 583–596.
- GRUNINGER, J., RATKOWSKI, A., AND HOKE, M. 2004. The sequential maximum angle convex cone (SMACC) endmember model. *Proc. SPIE, Algorithms for Multispectral and Hyper-spectral and Ultraspectral Imagery* 5425, 1–14.
- GUYON, I., WESTON, J., BARNHILL, S., AND VAPNIK, V. 2002. Gene selection for cancer classification using support vector machines. *Machine Learning* 46, 1–3, 389–422.
- HASTRUP, R. C., BELL, D. J., CESARONE, R. J., EDWARDS, C. D., ELY, T. A., GUINN, J. R., ROSELL, J. R., SRINIVASAN, J. M., AND TOWNES, S. A. 2003. Mars network for enabling low-cost missions. *Acta Astronautica* 52, 2–6, 227–235.
- IP, F., DOHM, J. M., BAKER, V. R., DOGGETT, T., DAVIES, A. G., CASTAÑO, R., CHIEN, S., CICHY, B., GREELEY, R., AND SHERWOOD, R. 2006. Development and testing of the autonomous spacecraft experiment (ASE) floodwater classifiers: Real-time smart reconnaissance of transient flooding. *Remote Sensing of Environment* 101, 463–481.
- KESHA, N. AND MUSTARD, J. 2002. Spectral unmixing. *IEEE Signal Processing Magazine* 19, 1, 44–57.
- MACQUEEN, J. B. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*. Vol. 1. University of California Press, Berkeley, CA, 281–297.
- MANDRAKE, L., WAGSTAFF, K. L., GLEESON, D., REBBAPRAGADA, R., TRAN, D., CASTAÑO, R., CHIEN, S., AND PAPPALARDO, R. 2009. Onboard detection of natural sulfur on a glacier via an svm and hyperion data. In *Proceedings of the IEEE Aerospace Conference*.
- MAZZONI, D., WAGSTAFF, K., BORNSTEIN, B., SCHARENBRUCH, L., TANG, B., AND RODEN, J. 2005. Pixellearn. JPL New Software Report NSR-42082.
- NASCIMENTO, J. AND DIAS, J. 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on* 43, 4, 898–910.
- PEARLMAN, J. S., BARRY, P. S., SEGAL, C. C., SHEPANSKI, J., BEISO, D., AND CARMAN, S. L. 2003. Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing* 41, 6, 1160–1173.
- PELKEY, S. M., MUSTARD, J. F., MURCHIE, S., CLANCY, R. T., WOLFF, M., SMITH, M., MILLIKEN, R., BIBRING, J.-P., GENDRIN, A., POULET, F., LANGEVIN, Y., AND GONDET, B. 2007. CRISM multispectral summary products: Parameterizing mineral diversity on Mars from reflectance. *Journal of Geophysical Research* 112, E08S14.
- RANGO, A., MARTINEC, J., CHANG, A. T. C., FOSTER, J. L., AND VAN KATWIJK, V. 1989. Average areal water equivalent of snow in a mountain basin using microwave and visible satellite data. *IEEE Transactions on Geoscience and Remote Sensing* 27, 6, 740–745.
- REBBAPRAGADA, U. AND BRODLEY, C. E. 2007. Class noise mitigation through instance weighting. In *Proceedings of the European Conference on Machine Learning*. Lecture Notes in Computer Science, vol. 4701. 708–715.
- REBBAPRAGADA, U., MANDRAKE, L., WAGSTAFF, K. L., GLEESON, D., CASTAÑO, R., CHIEN, S., AND BRODLEY, C. E. 2009. Improving onboard analysis of Hyperion images by filtering mislabeled training data examples. In *Proceedings of the IEEE Aerospace Conference*.
- SOTIN, C., HEAD, J. W., AND TOBIE, G. 2002. Europa: Tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting. *Geophysical Research Letters* 29, 8, 74.

- THOMPSON, D., MANDRAKE, L., GILMORE, M., AND CASTAÑO, R. 2010. Superpixel endmember detection. *IEEE Transactions on Geoscience and Remote Sensing* 48, 11, 4023–4033.
- TUCKER, C. J., DREGNE, H. E., AND NEWCOMB, W. W. 1991. Expansion and contraction of the Sahara desert from 1980 to 1990. *Science* 253, 5017, 299–300.
- UNGAR, S. G., PEARLMAN, J. S., MENDENHALL, J. A., AND REUTER, D. 2003. Overview of the Earth Observing One (EO-1) mission. *IEEE Transactions on Geoscience and Remote Sensing* 41, 6, 1149–1159.
- VAN RIJSBERGEN, C. J. 1979. *Information Retrieval*. Butterworth, London.
- WINTER, M. E. 1999. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Proceedings of the SPIE*. Vol. 3753. 266–275.